DOCUMENT RESUME

ED 163 049 TM 008 082

AUTHOR Herr, Gregory; Cohen Peter A.
TITLE Improving College Instruction: The Formative Assessment of College Teaching (FACT).
PUB DATE 30 Aug 78
NOTE 34p.; Paper presented at the Annual Meeting of the American Psychological Association (81st, Toronto, Ontario, Canada, August 30, 1978)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Effective Teaching; *Evaluation Criteria; *Formative Evaluation; Higher Education; *Rating Scales; *Student Evaluation of Teacher Performance; Teaching Assistants; Test Construction
IDENTIFIERS *Formative Assessment of College Teaching

ABSTRACT

It is evident that there is an existing need for a process-oriented assessment instrument which can be used to provide instructors with specific diagnostic feedback that can be utilized in the college classroom. The goal of the present study was to develop and validate a multidimensional student rating instrument which would provide such information. The Formative Assessment of College Teaching (FACT) was developed around assumptions which differentiate it from traditional student rating instruments. Eleven a priori scales formed the initial theoretical structure for the FACT. The field test version of the instrument contained 55 items which were rated on a five-point Likert-type scale. The FACT was completed by 777 undergraduate students enrolled in Introductory Psychology, who rated their graduate teaching assistant instructors. Data from the field test was cluster analyzed. Evidence confirming the existence of eight of the eleven a priori scales was found, and four response scales were also developed. The effects of response bias and of selected student background variables upon the 8 main scales of the revised, 24-item instrument were explored. Assumptions inherent in the use of the FACT for instructional improvement were made. (Author/JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED163049

IMPROVING COLLEGE INSTRUCTION: THE FORMATIVE ASSESSMENT OF COLLEGE TEACHING (FACT)

Gregory Herr and Peter A. Cohen

University of Michigan

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gregory Herr

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Presented at the Annual Meeting of the
American Psychological Association
Toronto, August 30, 1978

TM008 082

Student ratings of instructor effectiveness at the college level have become increasingly popular in recent years. Such ratings of faculty can be categorized either in terms of summative or formative evaluation efforts. When used in a summative fashion, ratings are often used as criteria for tenure, faculty promotion, and merit increases. Because of the possible career consequences of receiving low ratings, the rating procedure can be a source of anxiety for the instructor. In contrast, a formative evaluation of instructional performance is more diagnostic in nature with the intent of delivering specific potentially useful information to the instructor in a relatively non-threatening manner.

A recent development in the use of summative student ratings has been the implementation of "catalogs" of items from which the instructor can select items relevant to his/her particular classroom situation. Systems such as Purdue's CAFETERIA and the University of Michigan's Instructor Designed Questionnaire (IDQ) provide a number of dimensions from which items can be selected. While these catalog systems have become relatively sophisticated, they appear to be almost exclusively used in a summative fashion. If administrators and/or department chairpersons are to use course ratings for decision making, the instructor will certainly want to obtain favorable ratings. In this case the instructor's selection of items is likely to be influenced (i.e., selecting only those items for which favorable ratings are anticipated), and a loss of valuable diagnostic information may occur.

While the idea of mid-course assessment for the improvement of college instruction seems useful, research suggests that the feedback of mid-semester

---

The authors would like to express their appreciation to Jim Kulik, Chen-Lin Kulik, and Terry Davidson for their comments on earlier drafts of this paper.

ratings does not often lead to substantial change in instructor behavior as measured by end-of-semester ratings (Miller, 1971; Thomas, 1969). Centra (1973) found that while in general instructors did not tend to modify their teaching behaviors during the course after receiving mid-semester feedback, instructors who had unwarrantedly high opinions of their teaching did show positive change on some rating items. Pambookian (1974) discovered that the most change in behavior occurred in instructors who received a moderately favorable rating. Studies such as these imply that the effectiveness of mid-course evaluation may be moderated by the degree of discrepancy between student-instructor perceptions concerning important behavioral incidents in the classroom.

There are a number of reasons why feedback in the form of mid-course ratings might not be effective in facilitating positive change in instructor behavior. Kulik and McKeachie (1975) offer several explanations. First of all, perhaps more time than half a semester is required to produce change. It may be that students' initial impressions of the instructor are relatively stable and are resistant to change within a semester. Another explanation for the minimal amount of instructional modification is the absence of a set of norms concerning other classes. This explanation is based upon the assumption that an instructor can draw meaningful information from a course evaluation only by comparing his/her ratings with those of other instructors. As Centra and Linn (1976) point out, understanding the total context within which a particular course is taught may be the important factor in interpreting the meaning of student ratings. Kulik and McKeachie also suggest that instructors need to know <u>how</u> to change their teaching behavior in addition to receiving diagnostic feedback. In most studies, such consultative services are not provided. Finally, many rating instruments provide only the most global student impressions of the instructor along a set of general teaching dimensions. They do not provide much useful feedback targeted toward specific instructional behaviors.

The notion of a diagnostic rating instrument for use at the college level is not new. Cosgrove (1959) developed a diagnostic rating instrument almost twenty years ago. Ten sets of four adjective phrases yielded four separate scores which could be profiled. More recent efforts in developing diagnostic instruments have been reported by several investigators (Greenwood, Bridges, Ware, & McLean, 1973; Meredith, 1976; Polmann & Elmore, 1976). Olson (1978) developed a set of behavior-specific items to correspond with each of the general rating factors of the Student Instructional Rating System (SIRS) (Davis, 1969). Olson proposed that the SIRS instrument could be administered by instructors as a general survey in order to identify areas of special concern or deficiency. Once general areas have been identified, behavior-specific items can be used to elicit information from students in a form which is conducive to use by instructors. Inherent in this proposed procedure seem to be two assumptions. First, effective formative evaluation of instruction would seem to be at least a two-stage process. Secondly, separate instruments are needed to identify general areas for change and to identify specific instructor practices.

The purpose of the present research was to develop a multidimensional student rating instrument which could be used by instructors <u>during</u> a course to delineate areas for change. Although summative instruments such as CAFETERIA, IDQ, and SIRS could perhaps be used for this purpose with some success, it would make better sense to retain these instruments for use as outcome measures. A separate instrument for mid-course assessment would be especially valuable if there were numerous scales which could be profiled for high-impact feedback to instructors. In addition, the case for the validity of the instrument would be strengthened if scales for assessing the effects of bias due to response sets (Cronbach, 1946; Cronbach,1950) were built in. Such a measure might help

bridge the gap which presently exists between global ratings of instruction and behavior-specific instruments.

Method

A conceptual as well as an empirical framework guided the initial development of the instrument, the Formative Assessment of College Teaching (FACT). After a review of many summative instruments and informal discussions with college teachers, items to correspond with 11 a priori scales were written. The 11 scales and corresponding measurement intent for each are summarized in Table 1.

Insert Table 1 about here

The assumptions underlying item development were essentially those of scaling theory rather than of measurement theory. Although it is convenient for statistical reasons to select items which elicit a normal probability distribution of responses, such a selection criterion does not always make the best conceptual sense. Figure 1 represents a hypothetical frequency distribution of responses (N=20) for an imaginary rating item which meets the assumptions of interval measurement. Although the resulting normal distribution is psychometrically pleasing, it can be seen that responses in the "no opinion" category are maximized. Since "no opinion" has virtually no informational value from a diagnostic point of view, responses in this category are not particularly desirable. In addition, it should be noted that the remaining responses are, at least in theory, divided equally between two groups signifying agreement and disagreement with the content of the item. In contrast, Figure 2, represents a hypothetical item characterized by a markedly skewed distribution. Such a distribution results when respondents are in strong agreement about the

content of the item being rated. Responses in the "no opinion" category tend to be minimized but are not eliminated.

Insert Figures 1 and 2 about here

Subjects.--Subjects were graduate student teaching fellows and their undergraduate students enrolled in Introductory Psychology at the University of Michigan during the 1977-78 academic year. During the fall semester, approximately 340 students enrolled in 14 separate sections of the course took part. During the following semester, 437 students in 19 sections of the course participated. The graduate student teaching fellows had complete responsibility for teaching the sections of the course which were assigned to them.

When both samples are considered together, the total number of student participants was 777. The sample was almost evenly divided into males and females. Approximately 93% were between 18 and 21 years of age with a median age of 18.5 years. About 84% of the sample were white, 5% black, 8% other minority and the remainder unidentified. About 89% were freshman and sophomores; 9% were upperclass students; the remaining 2% were graduate students or unidentified. Sixty-four percent of the sample reported that they elected the course primarily because of interest while 22% reported that it was required of them.

Instrumentation.--The field test version of the FACT consisted of 55 items covering various aspects of the course and of the instructor's style and personal characteristics.[1] The respondents were asked to agree or disagree with each statement according to a five-point scale (5=strongly agree; 4=agree; 3=neutral/no opinion; 2=disagree; 1=strongly disagree). Six background items (sex, class, age, ethnic group, reason for taking the course, and college major) were included in a final section of the instrument. Because of the sensitive nature of the instructional rating task, students were asked not to

identify themselves. Respondents were assured that the instructor would not find out how any individual student responded to any question. They were also informed that the background information would not be used by the researchers to identify any student who participated in the rating task.

Procedure.--The first field test of the FACT was conducted at the University of Michigan during the seventh week of the Fall 1977 semester. Approximately 12 graduate student teaching fellows who were teaching one or more sections of Introductory Psychology were invited to participate. Prior to deciding whether or not to participate, instructors were given a copy of the instrument, and the proposed procedures were explained. Instructors were assured that they would receive feedback in the form of descriptive statistics for each item within one week of their use of the instrument. Potential users were also assured that only they and the researchers would have access to the summarized results of the rating information. Of the 12 instructors who were aware that the instrument was available, eight chose to make use of it in their classes. A second field test of the instrument was conducted during the seventh week of the Winter 1978 semester. Data from the second field test reflect the participation of 12 teaching fellows and their students.

During both field tests the instruments were distributed to students by the instructor during a regular class session. The rating forms were then completed by the students according to directions printed on the instrument. Completed FACT forms were collected, placed in an envelope, and returned by a student volunteer to a central location for processing and analysis. Approximately one week after administration, each instructor received a printout showing descriptive statistics for each item of the instrument.

Results

Descriptive Statistics.--Measures of central tendency and variability for each of the 55 rating items were computed. Comparison of the means and medians revealed a slight to moderate skewness for most items. Further inspection revealed that approximately 5 of the 55 rating items were normally distributed; 5 were were normally distributed with a tendency toward skewness; 15 were clearly skewed; and 30 were markedly skewed. This preliminary finding is in accordance with a scaling approach to development of the instrument.

Matrix Preparation.--Since the distributions of most items were skewed, the suitability of Pearson's r as a measure of association or similarity was questioned. Kendall's (1970) tau b was considered to be a more appropriate statistic for this application. In order to determine if there was a practical difference in the magnitude of Pearson r's and tau b's, both statistics were computed for a set of 22 purposively-selected pairs of items which ranged in strength of association from very low (r=.03) to moderately high (r=.73). These statistics are shown in Table 2.

Insert Table 2 about here

For items which were not strongly associated, there was little difference in the values of the two statistics. For items which were strongly associated, however, there was a marked difference between the two. As a result of this comparison, it was decided to prepare a matrix of tau b's rather than a matrix of Pearson r's for further use in the cluster analysis.[2]

Cluster Analysis.--The computer program used for the cluster analysis was an implementation of Johnson's (1967) maximum hierarchical method.[3] This algorithm operates upon a matrix of similarity measures and yields solutions

which are invariant under monotone transformations of the similarity data. Clustering proceeds from the initial cluster of two elements to a final solution in which all elements are brought into one large cluster. The algorithm shares several features of Shepard's (1962a and 1962b) and Kruskal's (1964) multidimensional scaling techniques but yields a cluster diagram which greatly assists the analyst in the inspection of clusters and construction of scales.

Development of the Main Scales.--Inspection of the cluster diagram revealed that eight of the 11 a priori scales could be identified as cohesive clusters. Clusters which corresponded to Cognitive Diversity, Favoritism, and Instructor Skill were identifiable but too weak to justify inclusion in the revision of the instrument. Three items were selected for inclusion in each of the eight scales. Final item selection was determined by two criteria. First, items had to make conceptual sense and had to have obvious face validity. Secondly, tau b intercorrelations for items within a scale had to exceed an arbitrary absolute value of 0.35. Prior to scale construction, some items dictated reverse scoring based on the wording of the item statement. Linear additive indices were then constructed according the general formula

$$\bar{X}_{ijk} = (X_{1jk} + X_{2jk} + X_{3jk})/3$$

where X represents the value of item i for person j in cluster k, and $\bar{X}$ represents the mean value of the three items comprising the scale.

Intercorrelations of items which were used to construct the eight scales are displayed in Table 3. In Table 4 descriptive statistics for each of the eight main scales of the revised FACT can be seen. Comparison of the mean and median values of the scales for all valid cases revealed little discrepancy between the two central tendency measures. Since the scales did not appear to be very skewed, they were treated with interval-level parametric statistics in subsequent analyses.

Insert Tables 3 and 4 about here

Development of the Response Scales.---To determine what proportion of the variance in the main scales was attributable to extreme response set (ERS) and other response-related characteristics of the rater, five response scales were constructed. The Positive Response Scale (PRS) consisted of the sum of the number of items for which the respondent chose the most positive alternative. Similarly, the Negative Response Scale (NRS) consisted of the sum of the number of items for which the respondent chose the most negative alternative. To identify the tendency of some respondents to refrain from agreeing or disagreeing with an item, the No Opinion Scale (NOS) was constructed. Each response of "neutral/no opinion" contributed one point toward an individual's score on this scale. The number of unanswered items made up the Missing Response Scale (MRS). Finally a Response Inconsistency Scale (RIS) was constructed to detect inconsistencies in responses to items which are highly intercorrelated. One point toward the RIS was given if there was a discrepancy of three or four points between two items within a scale. For each of these five scales, low scores are desirable; higher scores reflect the biasing influence of a response tendency. A maximum value of 24 may be obtained on each response scale. Means, medians, and standard deviations for the five response scales are shown in Table 5.

Insert Table 5 about here

Comparison of the means and medians reveals that the Positive Response Scale, Missing Response Scale, and No Opinion Scale are moderately skewed. As shown in Table 6, correlations among the response scales ranged from 0.45 to 0.86 in absolute value.

Insert Table 6 about here

Relationships Among Scales.--Zero-order Pearson correlation coefficients among the main scales were computed. These correlations varied in absolute value from a low of 0.042 between Work Quantity and Course Standards to a high of 0.349 between Instructor Motivation and Social Interaction.

Insert Table 7 about here

For purposes of comparison, fourth-order partial correlations among the main scales were also computed. By partialling out the variance contributed by four of the response scales,[4] a better estimate of the relationship among main scales could be obtained. The partial correlations are displayed in Table 8.

Insert Table 8 about here

The partial correlations range in absolute value from a low of 0.001 between Instructor Motivation and Social Interaction to a high of 0.218 between Instructor Motivation and Personal Contact. The magnitude of the relationships among approximately eight variables was dramatically reduced through use of the partial correlation technique.

A further effort was made through the use of multiple regression analysis to explore the relationships among main scales and response scales. A set of eight multiple regression analyses were performed. For each analysis, four response scales were used as independent variables and one main scale was used as the dependent variable. Thus the relationships of response scales to all main scales was systematically explored. Results of the analyses are displayed in Table 9.

Insert Table 9 about here

The R^2 values indicate that no main scale is unaffected by the response variables when taken as a linear additive set of predictors. The Work Quantity Scale was least affected by the response scales (R^2=.01); Instructor Motivation was greatly affected (R^2=.22).

Effects of Background Variables.--Relationships between all FACT scales and the five background variables were explored through use of eta^2 (Hays, 1973), which is appropriate for use with two variables when one is interval level and the other is nominal level. Eta^2, like r^2, is interpretable as a measure of the variance shared by two variables. The background variables taken separately account for a very small proportion of the variance of each scale as can be seen in Table 10.

Insert Table 10 about here

Discussion

The development of student instructional rating instruments has traditionally proceeded under the assumptions of interval-level measurement. Initial development efforts on the FACT showed that scaling theory provides a sounder approach when rating items are highly skewed. One drawback of this approach, however, is that statistics for estimating the reliability of ordinal items are not readily available. In addition, statistics such as Cronbach's alpha probably provide greatly distorted estimates of internal consistency. This is one problem which will have to be dealt with in future research.

Comparison of zero-order and partial correlation matrices revealed marked differences in the magnitude of correlations for several main scales. For

instance, when the variance attributable to four response scales was partialled out, the correlation between Instructor Motivation and Social Interaction dropped from .349 to .001. The importance of the influence of the response scales is further highlighted by the surprisingly large multiple correlation of .47 obtained by regressing Instructor Motivation on four of the five response scales. Apparently extreme response set (ERS) (Borgatta & Glass, 1961) as measured by the Positive Response and Negative Response scales is a source of significant bias in the FACT. Although ERS has not been examined in most summative rating scales, it is probable that it is a potent source of bias in other student rating instruments as well. The results of many instructional rating forms may be quite heavily contaminated by the effects associated with the response style of the student. The implication of this finding is not that student ratings are hopelessly biased, but rather that the biasing influences should be statistically estimated and removed from ratings so that adjusted values can be used by instructors.

The influence of the five background variables is somewhat less than expected. Numerous studies have included similar variables, and the results seem to be inconsistent. No doubt the wide variety of instruments, institutions, and student populations have much to do with this inconsistency. In the present study the five variables appear to account for roughly 4% to 6% of the variance in any scale. The effects of background variables are small indeed in comparison to the effects of ERS and other variables which reflect the response style of the respondent.

As mentioned earlier, mid-semester rating schemes have not as yet proved useful as a means of directing change in instructor classroom behavior. The Formative Assessment of College Teaching (FACT) was developed to ameliorate some of the difficulties which have hindered mid-course evaluation efforts.

Four assumptions are integral to its utility as an instrument for instructional improvement.

First, while ratings used for summative purposes provide the instructor with an assessment of his/her overall teaching effectiveness along a number of general dimensions they are rarely useful in providing the kind of specific diagnostic information necessary for instructional improvement. The FACT, on the other hand, is designed around a class of potentially important instruction-related behaviors.

A second assumption is that utilization of instructional feedback will occur optimally under non-threatening conditions. Because the use of norms is not emphasized, the instructor need not be intimidated by inter-instructor comparisons. When an instructor chooses to use the FACT, only s/he and those involved in instructional consultation have access to the ratings. Immediate feedback of the results of the ratings also contributes to the climate of facilitating teaching rather than evaluating it.

Third, under such non-threatening conditions the instructor will seek information of a change-oriented rather than of an ego-building nature. It is assumed that discrepancies between student ratings and instructor self-perceptions concerning specific teaching styles and behaviors will motivate instructors to modify their teaching practices in directions they consider to represent quality instruction. No assumptions concerning a universal definition of good teaching are made.

A fourth assumption concerning the FACT is that a formative assessment of teaching should be geared toward utilization of the results. Information feedback is not equivalent to utilization. It is not assumed that teachers will become more effective simply by acquiring information on their teaching via the FACT. Rather, they must know additionally how to become more effective

teachers. Therefore, the FACT is only part of what should be considered a process for the ongoing development of more effective instruction.

The value of diagnostic rating systems might be more apparent if instructors were observed to make changes in instruction-related behavior after receiving diagnostic feedback. Perhaps it is not the mere presence or absence of feedback which makes a difference--the quantity, timing, format, and quality of feedback are four of many variables which should also be explored. Experimental designs in which instructors receive feedback of differing types would seem to be useful in answering some basic questions concerning the effectiveness of feedback conditions in moderating change. In cases where it is not possible to manipulate variables, perhaps causal modelling can be used to gain insight into the process of change. Further research focusing on process-oriented feedback mechanisms is needed if the efficacy of student ratings for course improvement is to be adequately demonstrated.

Footnotes

1. A copy of the field test version of the instrument may be obtained from the first author.
2. A copy of the tau b matrix (formatted as an OSIRIS type 2 data matrix) may be obtained from the first author.
3. The cluster analysis program used was HICLUSTER, which is contained in the OSIRIS III collection of computer programs developed at the Survey Research Center, University of Michigan.
4. Since it was impossible to compute values for the MRS when listwise deletion of cases with missing data was chosen as an option of analysis, the MRS was not included in this and in other analyses involving the response scales.

Reference Notes

Davis, R. H. Student Instructional Rating System (SIRS) (Tech. Rep.). East Lansing, Michigan: Michigan State University, Office of Evaluation Services, December, 1969.

Olson, L. A. Behavior-specific items for student rating of instruction. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, March, 1978.

References

Borgatta, E. F. & Glass, D. E. Personality concomitants of extreme response sets (ERS). Journal of Social Psychology, 1961, 55, 213-221.

Centra, J. A. Effectiveness of student feedback in modifying college instruction. Journal of Educational Psychology, 1973, 65, 395-401.

Centra, J. A. & Linn, R. L. Student points of view in ratings of college instruction. Educational and Psychological Measurement, 1976, 36, 693-703.

Cosgrove, D. J. Diagnostic rating of teacher performance. Journal of Educational Psychology, 1959, 50, 200-204.

Cronbach, L. J. Further evidence on response sets and test designs. Educational and Psychological Measurement, 1950, 10, 3-31.

Cronbach, L. J. Response sets and test validity. Educational and Psychological Measurement, 1946, 6, 475-494.

Greenwood, G. E., Bridges, C. M., Jr., Ware, W. B., & McLean, J. E. Student evaluation of teaching behaviors instrument: A factor analysis. Journal of Higher Education, 1973, 44, 596-604.

Hays, W. L. Statistics for the Social Sciences (2nd ed.). New York: Holt, Rinehart, and Winston, 1973.

Johnson, S. C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 241-254.

Kendall, M. G. Rank Correlation Methods (4th ed.). London: Griffin, 1970.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to nonmetric hypothesis. Psychometrika, 1964, 29, 1-27.

Kulik, J. A. & McKeachie, W. J. The evaluation of teachers in higher education. In F. N. Kerlinger (Ed.) Review of Research in Education, Vol. 3, Itaska, IL: F. E. Peacock Publishers, Inc. 1975.

Meredith, G. M. Structure of the teaching analysis by students (TABS). Journal of Psychology, 1976, 92, 235-242.

Miller, M. T. Instructor attitudes toward, and their use of, student ratings of teachers. Journal of Educational Psychology, 1971, 62, 235-239.

Pambookian, H. S. The initial level of student evaluation of instruction as a source of influence on instructor change after feedback. Journal of Educational Psychology, 1974, 66, 52-56.

Pohlman, J. T. & Elmore, P. B. The instructional improvement questionnaire. Journal of Educational Measurement, 1976, 13, 161-163.

Shepard, R. N. Analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika, 1962a, 27, 125-140.

Shepard, R. N. Analysis of proximities: Multidimensional scaling with an unknown distance function. II. Psychometrika, 1962b, 27, 219-246.

Thomas, H. B., Jr. Changing teacher behavior: An experiment in feedback from students to teachers. Dissertation Abstracts International, 1969, 5-A, 1903-1904.

TABLE 1

Names of A Priori Scales and
General Measurement Intent for Each Scale

Scale Name	Measurement Intent
Cognitive Diversity	Student's perception of the instructor's degree of support for widely varying opinions and ideas
Course Sequencing	Student's perception of the degree of sequential, progressive course organization
Course Standards	Student's perception of the standards of performance set by the instructor.
Favoritism	Student's perception of the degree to which the instructor pays attention to some students in the class more than others
Instructor Motivation	Student's perception of the instructor's level of enthusiasm in teaching
Instructor Skill	Student's perception of the instructor's ability to communicate information to students
Personal Contact	Student's perception of the instructor's efforts to establish a personal relationship with the student
Risk Taking Climate	Student's perception of the classroom climate in respect to freedom from threat or derision by the instructor
Social Interaction	Student's perception of the degree to which the instructor fosters the development of interpersonal relationships among class members
Student Centeredness	Student's perception of the "student vs. teacher centered" dimension of the course
Work Quantity	Student's perception of the reasonableness of the quantity of work required in the course

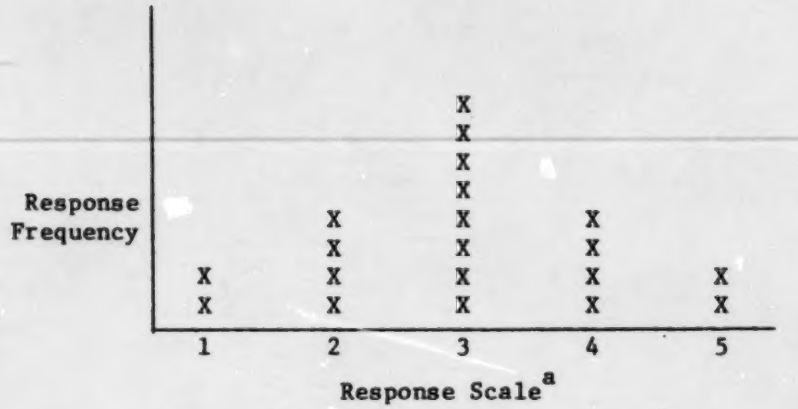

Figure 1. Hypothetical Frequency Distribution of a Rating Item Selected Under the Constraints of Interval Measurement (N=20).

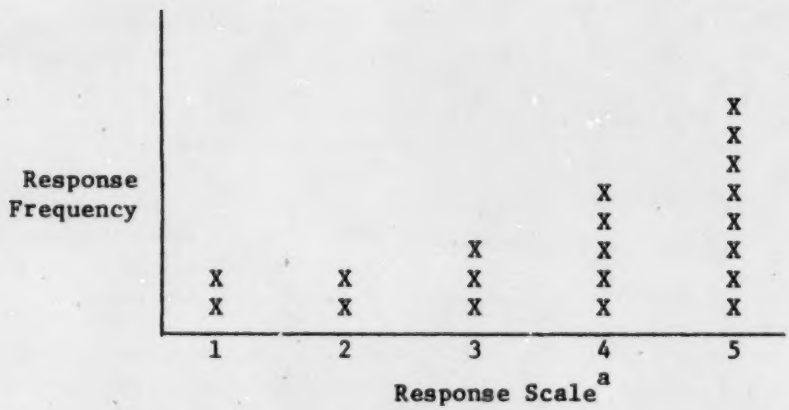

Figure 2. Hypothetical Frequence Distribution of a Rating Item Selected Under the Constraints of Ordinal Scaling (N=20).

[a] Response scale values are 1=strongly disagree, 2=disagree, 3=no opinion, 4=agree, 5=strongly agree

Table 2

Comparison of the Magnitude of Pearson's r and Kendall's tau b for Selected Items

	Statistic				Statistic		
Item Pair	N	Pearson r	Tau b	Item Pair	N	Pearson r	Tau b
7,24	767	.0274	.0432	11,27	765	-.3961	-.3612
5,13	767	-.0773	-.0779	11,31	757	.4241	.3705
4,37	755	.0817	.0909	16,42	760	.4770	.3861
5,28	767	-.0884	-.0649	14,40	764	.4790	.4306
20,39	768	.1600	.2128	34,49	752	-.5316	-.4550
10,13	766	.1818	.1924	40,46	761	.5660	.5243
15,52	758	-.1962	-.2021	26,44	758	-.5762	-.5283
19,38	758	.2693	.2494	41,43	752	.6047	.5459
29,35	763	.3144	.3469	31,55	752	.6465	.5613
27,36	759	-.3235	-.2936	42,50	759	.6996	.5930
45,48	761	.3274	.3229	36,44	753	.7299	.6588

Note: Based upon analysis of entire sample (N=777). Figure in column labeled N represents the actual number of cases per item pair.

Table 3

Item Intercorrelations for Each Scale in the Revised Version of the FACT (Form B1)

Scale Name	Intercorrelations	
	Item Pair	Tau b
Work Quantity	1,28	-.538
	1,53	-.438
	28,53	.434
Course Sequencing	3,13	-.571
	3,31	.528
	13,31	-.502
Course Standards	9,25	.423
	9,49	.461
	25,49	.624
Instructor Motivation	15,42	.566
	15,50	.541
	42,50	.593
Risk Taking Climate	20,38	-.390
	20,51	.392
	38,51	-.491
Student Centeredness	5,41	.473
	5,43	-.420
	41,43	-.546
Personal Contact	8,18	-.585
	8,26	-.544
	18,26	.613
Social Interaction	22,47	.465
	22,54	.559
	47,54	.605

Table 4

Descriptive Statistics for the Eight Main Scales

Scale Name	Median	Mean	Standard Deviation	N
Work Quantity	2.82	2.88	0.47	752
Course Sequencing	3.14	3.07	0.52	756
Course Standards	3.57	4.05	0.99	752
Instructor Motivation	3.97	3.83	0.89	757
Risk Taking Climate	2.45	2.50	0.41	751
Student Centeredness	3.09	3.02	0.56	758
Personal Contact	2.51	2.60	0.46	764
Social Interaction	4.01	3.88	0.87	751

Table 5

Descriptive Statistics for the Five Response Scales

Scale Name	Median	Mean	Standard Deviation
Positive Response	3.67	4.46	3.59
Negative Response	3.96	3.64	2.37
Missing Response	0.04	0.24	1.50
Response Inconsistency	4.62	4.65	2.85
No Opinion	3.58	4.08	3.15

Note: Based upon 777 cases

Table 6

Matrix of Pearson Correlations Among Four Response Scales

	Negative Response Scale	Positive Response Scale	No Opinion Scale
Response Inconsistency Scale	.749	.756	-.601
Negative Response Scale		.553	-.451
Positive Response Scale			-.525

Note: Based upon 695 complete cases. Since cases with missing data were deleted from the analysis, the Missing Response Scale could not be included here. With 693 degrees of freedom, all correlations are significant beyond the .001 level.

Table 7

Matrix of Zero-Order Pearson Correlations Among the Main Scales[a]

Scale Name	Course Sequencing	Course Standards	Instructor Motivation	Risk Taking Climate	Student Centeredness	Personal Contact	Social Interaction
Work Quantity	-.064	-.042	-.053	.138**	-.051	.051	-.093*
Course Sequencing		.330**	.233**	.036	.100**	-.048	.150**
Course Standards			.303**	-.064	.162**	-.126**	.186**
Instructor Motivation				-.057	.298**	-.241**	.349**
Risk Taking Climate					-.034	.102**	-.095**
Student Centeredness						-.078**	.162**
Personal Contact							-.252**

[a] Correlation coefficients are based upon 695 cases with no missing data.

* $p < .05$
** $p < .01$

Table 8

Partial Correlations Among Main Scales with the Variance Contributed by Four Response Scales Removed

	Course Sequencing	Course Standards	Instructor Motivation	Risk Taking Climate	Student Centeredness	Personal Contact	Social Interaction
Work Quantity	-.133**	-.107**	-.084*	.017	-.083*	.033	-.102**
Course Sequencing		.102**	-.074	-.042	-.093*	-.043	-.119**
Course Standards			-.035	-.171**	-.047	-.130*	-.121**
Instructor Motivation				-.094*	-.094*	-.218**	-.001
Risk Taking Climate					-.081*	-.038	-.072
Student Centeredness						-.060	-.042
Personal Contact							-.174**

[a] Positive Response, Negative Response, No Opinion, and Response Inconsistency scales.

Note: Correlations are based upon 695 cases with no missing data.

* $p < .05$
* $p < .01$

Table 9

Regression Analysis of 8 FACT Scales with 4 Response Scales[a] as Predictors

Name of Criterion Variable	Standardized Beta Weights of Predictors: Response Inconsistency Scale	Negative Response Scale	Positive Response Scale	No Opinion Scale	R	R^2
Work Quantity	-.09	.00	-.03	.01	.11	.01
Course Sequencing	.02	-.22	.36	-.03	.33***	.11
Course Standards	.33	-.22	.17	-.01	.33***	.11
Instructor Motivation	.42	-.22	.23	-.01	.47***	.22
Risk Taking Climate	-.06	-.17	-.03	.03	.25***	.06
Student Centeredness	-.01	-.06	.23	-.04	.23***	.05
Personal Conduct	-.01	.03	-.21	.08	.25***	.06
Social Interaction	.17	.01	.22	-.07	.42***	.18

[a] All analyses are based upon 695 cases with no missing data.

*** $p < .001$

Table 10

Relationships Among FACT Scales and Five Student Background Variables

Scale Name	Background Variables									
	Sex		Class		Age[a]		Ethnic Group		Course Motivation[b]	
	Eta	Eta^2	Eta	Eta^2	Eta	Eta^2	Eta	Eta^2	Eta	Eta^2
Work Quantity	.056	.003	.085	.007	.152	.023	.031	.001	.071	.005
Course Sequencing	.038	.001	.083	.007	.120	.014	.036	.001	.098	.010
Course Standards	.033	.001	.109	.012	.135	.018	.013	.000	.081	.007
Instructor Motivation	.061	.004	.056	.003	.140	.020	.010	.000	.089	.008
Risk Taking Climate	.089	.008	.069	.005	.171	.029	.035	.001	.088	.008
Student Centeredness	.064	.004	.058	.003	.108	.012	.009	.000	.093	.009
Personal Contact	.005	.000	.057	.003	.115	.013	.007	.000	.086	.007
Social Interaction	.038	.001	.158	.025	.131	.017	.037	.001	.083	.007
Positive Response	.098	.010	.115	.013	.129	.017	.017	.000	.077	.006
Negative Response	.168	.028	.118	.014	.127	.016	.001	.000	.057	.009
Missing Response	.029	.001	.080	.006	.078	.006	.031	.001	.093	.009
Response Consistency	.086	.007	.079	.006	.117	.014	.027	.001	.065	.004
No Opinion	.041	.002	.104	.011	.113	.013	.070	.005	.124	.015

[a] Since age is an interval level variable, eta and eta^2 slightly overestimate the magnitude of the relationship reported.

[b] Reason for electing the course: interest vs. requirement.